
知识管理论坛
2016 年 第 3 期(总第 3 期)

【学术探索】

垂直知识图谱的构建与应用研究

◎ 阮彤[1] 王梦婕[2] 王昊奋[1] 胡芳槐[3]

[1] 华东理工大学信息学院 上海 200237
[2] 华东理工大学计算机科学与工程系 上海 200237
[3] 上海海翼知信息科技有限公司 上海 200433

摘要：[目的 / 意义] 近年来，知识图谱技术受到学术界和工业界的普遍关注。提出数据驱动的增量式知识图谱构建方法，为构建垂直知识图谱提供一种新思路。同时，通过 3 个用例研究提供垂直知识图谱的应用示范。[方法 / 过程] 首先给出知识图谱的形式化定义，然后提出数据驱动的增量式知识图谱构建方法，重点研究构建垂直知识图谱数据图的细节与难点。基于该方法，本文构建了中医药知识图谱、海洋知识图谱和企业知识图谱。[结果 / 结论] 以上垂直知识图谱的构建证实了本方法的可行性，它们各自的垂直应用体现了知识图谱的广泛应用。

关键词：知识获取 知识融合 语义搜索 辅助开方 关系发现

分类号：TP391

引用格式：阮彤 , 王梦婕 , 王昊奋 , 等. 垂直知识图谱的构建与应用研究 [J/OL]. 知识管理论坛 , 2016, 1(3): 226-234[引用日期]. http://www.kmf.ac.cn/paperView?id=43.

1 引言

自从语义网络的概念提出以来，大量的链接开放数据（Linked Open Data，简称 LOD）和用户生成内容（User-generated Content，简称 UGC）发布在互联网中，互联网从仅包含网页与网页之间超链接的文档万维网逐步转变为包含大量描述实体和实体之间丰富关系的数据万维网。在此背景下，为改善搜索引擎效果，谷歌公司于 2012 年提出“知识图谱”的概念[1]：一种描述真实世界客观存在的实体、概念及它们之间的关联关系的语义网络。

基于知识图谱的应用领域，本文将知识图谱分为通用知识图谱和垂直知识图谱（或行业知识图谱）。通用知识图谱不面向特定领域，可将其类比为“结构化的百科知识”。这类知识图谱包含了大量常识性知识，强调知识的广度。具有代表性的大规模通用知识图谱有 YAGO[2]、DBpedia[3]、Freebase[4]、NELL[5] 等，中文通用知识图谱有 Zhishi.me[6] 和 SSCO[7]。垂直知识图

基金项目：本文系国家高技术研究发展计划“心血管疾病与肿瘤疾病中西医临床大数据处理分析与应用研究”(项目编号：2015AA020107）研究成果之一。

作者简介：阮彤（ORCID：0000-0002-3546-8338），自然语言处理与大数据研究室主任，中国计算机学会大数据专委会常务委员，中文信息处理学会 CCIR 专委会委员，副教授，博士，ruantong@ecust.edu.cn；王梦婕（ORCID：0000-0003-4068-8869），硕士研究生；王昊奋（ORCID：0000-0002-0672-8081），自然语言处理与大数据研究室副主任，中国计算机学会术语工作委员会执委，中文信息学会语言与知识计算专委会委员，讲师，博士；胡芳槐（ORCID：0000-0002-7747-6686），总裁，博士。

收稿日期：2016-03-10 发表日期：2016-06-26 本文责任编辑：王铮

知识管理论坛，2016(3):226–234
DOI：10.13266/j.issn.2095–5472.2016.027

谱则面向特定领域，基于行业数据构建，强调知识的深度。垂直知识图谱可以看作基于语义技术的行业知识库，其潜在使用者是行业的专业人员。

在通用知识图谱的构建方面，已有相对成熟的技术和知识图谱产品，例如各大搜索引擎公司发布的谷歌知识图谱、百度“知心”、搜狗“知立方”等商用知识图谱。而在垂直知识图谱的构建方面，现有垂直知识图谱常采用手工构建方式，缺乏一套统一的垂直知识图谱构建方法。基于此，本文面向垂直知识图谱，首先对其进行形式化定义，然后提出数据驱动的增量式知识图谱构建方法：从多种类型的数据源出发，研究知识获取、融合过程中的细节与难点。最后，本文利用所提出的知识图谱构建方法构建了中医药知识图谱、海洋知识图谱和企业知识图谱，并对各自的垂直应用加以阐述，证实了本文方法的可行性和垂直知识图谱的广泛应用性。

2 知识图谱的形式化定义

通用知识图谱与垂直知识图谱的本质并无区别，因此本文对两类知识图谱统一地进行定义。如图 1 所示，知识图谱 G 由模式图 G_s、数据图 G_d 及二者之间的关系 R 组成，即 G=<G_s, G_s, R>。模式图 G_s=<N_s, E_s>，其中 N_s 表示类结点的集合，E_s 表示属性边的集合。模式图 G_s 中的类（结点）即为知识图谱中的概念，而属性（边）则对应概念之间的语义关系，包括 rdfs:subClassOf、rdfs:equivalent Class 这类来自语义网络现有标准 RDFS[8] 的属性和 employer 等用户自定义的属性。与此类似，数据图 G_d=<N_d, E_d> 中的结点集包含实例结点和字符串结点，边集合 E_d 中的边连接两个结点表示一条三元组事实，如 <Bill_Gates, alaMater, Harvard_University>。此处，实例即实体，表示计算机可识别的客观世界对象，而字符串常作为实例的某一属性值。模式图 G_s 和数据图 G_d 之间的关系 R 由 rdf:type 构成，表示数据图中的实例与所属概念之间的关系。

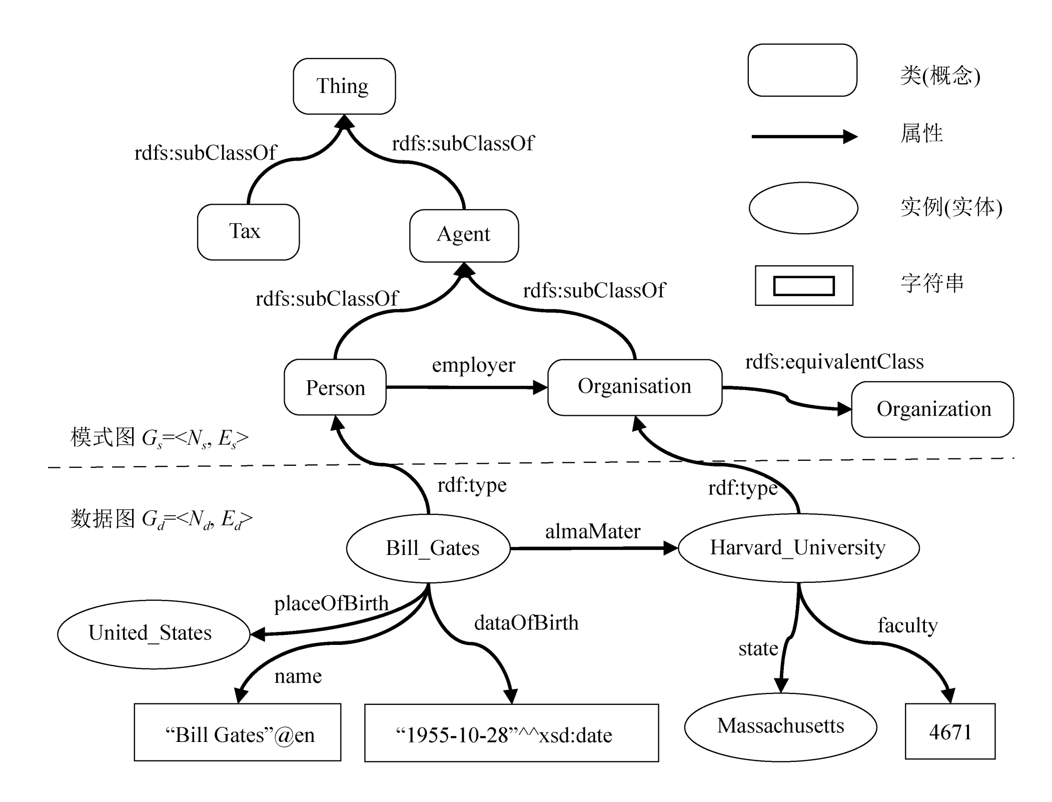


图 1 知识图谱的定义示例

知识图谱具有多方面技术优势：首先，知识图谱易于修改数据模式，具有良好的动态可扩充性。在构建知识图谱时可以利用该特性进行增量式的数据模式设计。其次，知识图谱的语义互操作特性和“链接数据”原则，使得不同来源的数据集成更为方便。此外，知识图谱支持 RDFS、OWL[9]、SPARQL[10] 等现有标准，可以逐渐要求内容供应商提供支持。最后，知识图谱显式地表达实体之间的关系，可用于开发语义检索、自动问答等应用。

3 相关工作

在知识图谱的构建方面，已经积累了大量通用知识图谱构建的工作。早期主要通过人工构建的方式，形成了 WordNet[11]、ResearchCyc[12] 等通用知识图谱。此后，大量知识图谱基于维基百科进行构建，如 YAGO、DBpedia 等。但由于抽取的目标数据不同，它们的知识丰富度各有差异 [13]。其中，DBpedia 抽取了维基百科中信息框中的所有信息和统计信息；而 YAGO 仅从维基百科中抽取其自定义的属性，并使用 WordNet 进行数据整合，因而准确率更高，但知识丰富度低于 DBpedia。不同于上述工具，Zhishi.me 和 SSCO 专注于构建中文知识图谱，除了使用中文维基百科，还额外使用互动百科和百度百科这两个非常流行的中文百科站点。近年来，基于开放域知识抽取的知识图谱构建项目受到关注，如 KnowItAll[14]、NELL。它们使用增量迭代的方法从大量的网页数据中学习出高质量三元组来构建知识图谱。

然而，由于垂直知识图谱与通用知识图谱的应用范围不同，它们采取的构建方法也有所区别。上述通用知识图谱采取自底向上的方式进行构建，这种方法有利于发现新的知识图谱模式。而垂直知识图谱注重知识的层次结构，通常需要预先构建模式图。由于通用知识图谱的构建方法不适用于垂直知识图谱，而现有的高质量垂直知识图谱常采用手工构建的方法，本文提出了数据驱动的增量式知识图谱构建方法，为自动地构建垂直知识图谱提供一套通用的方法。本文通过自顶向下的方式构建知识图谱的模式图，自底向上的方式构建数据图。这种方式可以保障数据抽取的质量。在具体的构建过程中，本文借鉴了已有的通用知识图谱构建方法：将百科知识作为一类重要的知识源，同时将增量迭代方法用于文本类型的知识抽取。

4 垂直知识图谱的构建

4.1 总体流程

由于垂直知识图谱强调知识的深度和整体的层次结构，在构建时常采用自顶向下和自底向上相结合的方式。其中，自顶向下的方式是指通过本体编辑器或手工构建的方法预先构建垂直知识图谱的模式图，进而构建数据图。而自底向上的方式指在构建数据图时，利用多种抽取技术获得知识源中的实体、属性和关系，并将这些置信度高的抽取结果合并到知识图谱中。

正如图 1 所示，知识图谱 G 由模式图 G_s、数据图 G_d 及二者之间的关系 R 组成。本文在已经构建了垂直知识图谱模式图 G_s 的前提下，从数据源出发，采用自底向上的方式说明构建垂直知识图谱数据图 G_d 和关系 R 的过程。

如图 2 所示，知识来源主要分为结构化知识、半结构化知识和非结构化知识。对于结构化知识，有大量的链接开放数据和存放在关系数据库中的领域知识。对于半结构化知识，维基百科、互动百科、百度百科等百科网站提供的信息框（Infobox）是一种半结构化知识。此外，不同领域下的垂直站点包含了大量的表格、列表数据，这也是半结构化知识。非结构化知识是指网络数据中大量的纯文本内容，其知识覆盖度最广，但抽取难度也最大。

从知识来源出发，主要通过知识获取和知识融合两个步骤构建知识图谱。根据知识图谱本身的特性，我们可以使用增量迭代的方式不断丰富所构建的知识图谱。这一构建过程称为数据驱动的增量式知识图谱构建。

知识管理论坛，2016(3):226–234
DOI：10.13266/j.issn.2095–5472.2016.027

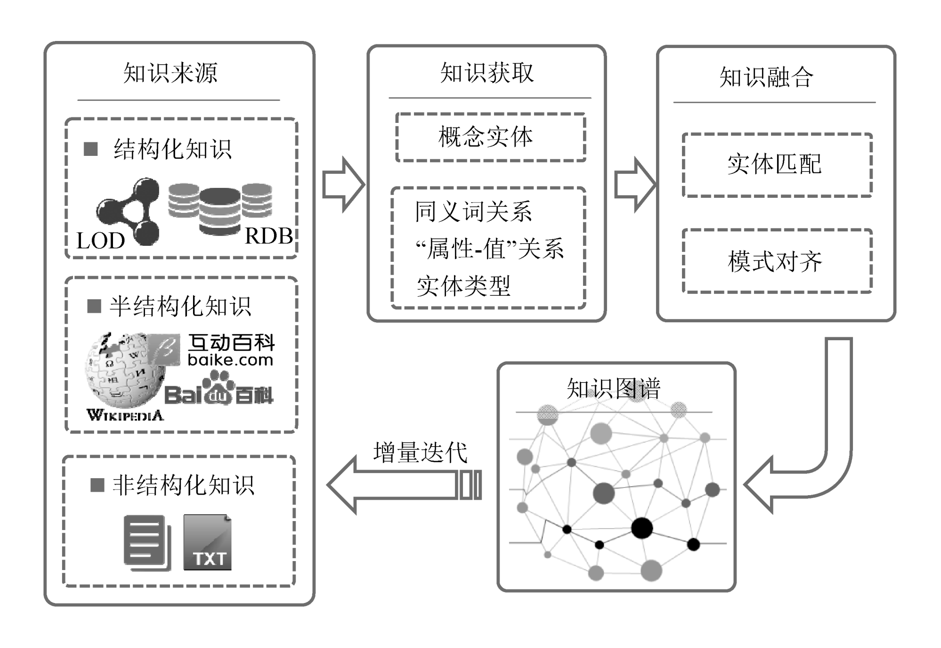


图 2 数据驱动的增量式知识图谱构建

4.2 知识获取

知识获取阶段需要从知识源中获取实体、同义词关系、"属性-值"关系以构建数据图 G_d，同时需要获取实体类型以构建关系 R。由于知识来源众多，且不同知识源之间存在数据重合，因此如何针对不同的知识源类型采用合适的抽取方法，并充分利用知识源之间的数据冗余性是知识获取阶段的难点。

本文作者提出多策略学习的方法进行知识获取[15]。多策略学习是指利用不同知识源之间的冗余信息，使用较易抽取的信息来辅助抽取那些不易抽取的信息。结构化知识和半结构化知识由于具有显式的结构和固定的格式，属于易抽取的信息，而无结构的文本知识属于较难抽取的信息。如图 3 所示，对于结构化知识中的关系数据库数据，可以通过 D2R（Relational Database to RDF）映射的方法将其转化成知识图谱中的链接数据。对于百科数据中的信息框、表格等半结构化知识，使用基于封装器(Wrapper)的抽取方法。封装器是面向某一具有特殊结构的数据源的信息抽取方法。对以上两类知识进行抽取，并将抽取结果加入种子集中。

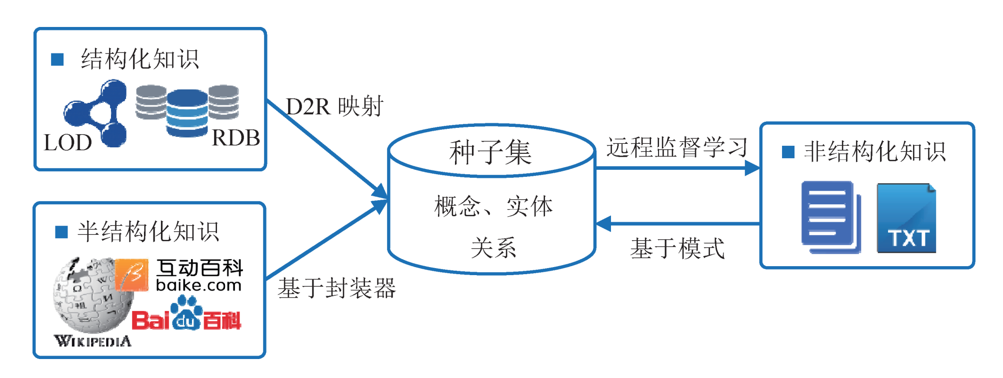


图 3 多数据源的抽取示意

对于无结构的纯文本知识，采用远程监督（Distant Supervision）[16] 和基于模式的方法相结合的增量迭代抽取方式。远程监督是一种基于假设“如果两个实体存在某种关系，那么任何包含这对实体的句子都很有可能表达相同的关系”、利用已知的实体关系对自动标注文本的方法。这里就可以利用种子集自动标注文本数据，然后根据标注结果自动地生成高质量的模式。利用这些模式到文本中学习新的知识，并加入到种子集中。这一过程不断迭代，直至没有新的知识被学习出来。

4.3 知识融合

知识获取阶段仅仅是从不同类型的知识源抽取构建知识图谱所需的实体、属性和关系，形成了一个个孤立的抽取图谱。为了形成一个完整的知识图谱，需要将这些抽取结果集成到知识图谱中，以进行知识融合。在进行知识融合时，需要解决多种类型的数据冲突问题，包括一个短语对应多个实体、实体属性名不一致、实体属性缺失、实体属性值不一致、实体属性值一对多映射等情况。知识融合阶段主要对数据进行实体匹配和模式对齐。

实体匹配旨在发现具有不同标识但代表真实世界中同一对象的那些实体，并将这些实体合并为一个具有全局唯一标识的实体对象添加到知识图谱中。目前常采用聚类的方法进行实体匹配，其关键在于定义合适的相似度度量。这些相似度度量常参考实体的以下特征：①字符相似：具有相同描述的实体可能代表同一实体；②属性相似：具有相同属性 - 值关系的实体可能代表同一对象；③结构相似：具有相同的相邻实体可能指向同一个对象。

模式对齐主要包括实体属性和属性值的整合。对于实体属性的整合，可以考虑的特征有属性的同义词、属性两端的实体类型，以及属性在抽取过程中对应的模式等。当融合来自不同知识源的数据出现数据冲突时，还可以考虑知识源的可靠性以及不同信息在各知识源中出现的频度等因素。本文作者对搜索引擎提供的知识卡片进行合并[17]，提供了一种在线知识融合的思路。该方案首先提出一种基于概率的实体评分算法找与知识卡片最相关的维基百科词条，由此合并代表同一实体的不同知识卡片。然后，将维基百科的信息框与 DBpedia 本体的映射关系作为训练数据，设计四维特征训练出属性对齐模型。最后使用相似度阈值对属性值进行去重与合并，形成值簇。

5 垂直知识图谱的用例研究

本文利用数据驱动的增量式知识图谱构建方法分别构建了中医药知识图谱、海洋知识图谱和行业知识图谱。下文将分别阐述这 3 个垂直知识图谱的构建过程和具体应用，以说明本文方法的有效性和垂直知识图谱的广泛应用。

5.1 中医药知识图谱

中医药领域已经积累了大量专业知识的分类信息，我们可以根据这些知识构建中医药知识图谱的模式图。目前主要基于中医病证分类与代码（国家标准）、中华中医药学会提供的诊疗指南、上海中医药大学附属曙光医院的药品数据库构建了中医药知识图谱的疾病库、证库等子库的模式图。对于中医药知识图谱数据图的构建，本文分别使用 D2R 映射方法从曙光医院的关系数据库中抽取药品信息；构造 Microsoft Office 软件的封装器，从“98 版证名分类标准”等国家标准以及曙光医院以 Microsoft Word 格式存储的临床知识库中抽取疾病、药方等信息；利用模式和远程监督结合的方法迭代地学习百科和中医药网站下的纯文本知识。由于从多个数据源中抽取数据，不同的数据源之间会存在重复或冲突。本文对数据源的可信度进行评分，基于数据来源以及数据在不同来源中出现的次数，对数据项进行排序，以解决数据冲突问题。

本文形成的中医药知识图谱主要包括疾病库、证库、症状库、中草药库和方剂库。基于中医药知识图谱可以进行中医药相关的自然语

知识管理论坛，2016(3):226–234
DOI：10.13266/j.issn.2095–5472.2016.027

言问答。同时，利用推理引擎 Drools[18]，可进行中医药辅助开方。

中医药方分为基础药方和经验药方。基础药方由疾病和证决定，经验药方则需要根据病人所患症状确定。当医生诊断出病人患有的疾病、证及症状后，经过中医药知识图谱的推理即可得到推荐药方。如图 4 所示，中医药知识图谱中存储的事实包含肝郁气滞证的基础药方和经验药方。推理引擎 Drools 将肝郁气滞证的药方转化为一系列规则。当输入的病人患有胁痛并被诊断为肝郁气滞证，根据规则只要使用基本方即可。但是，对于同时患有“口苦口干”症状的病人，根据规则，还需要去除川芎，增加牡丹皮等中草药。



图 4 中医药知识图谱用于辅助开方

5.2 海洋知识图谱

海洋知识图谱主要包括鱼类知识、海洋经济知识和海岛知识。其中，海洋经济知识由领域专家收集并存储在 Microsoft Word 文档中，本文使用 Microsoft Word 封装器将其转化映射成海洋知识子图。海岛知识源于舟山海洋数字图书馆提供的关系数据库，使用 D2R 映射工具 D2RQ[19] 完成数据转化，形成海岛知识子图。

对于鱼类知识，数据源众多，包括三大中文百科站点、台湾鱼类资料库（fishdb[20]）、世界鱼类分类阶层树状名录（FishBase[21]）、心食谱等行业站点，以及《中国食物成分表》(2002 年版)等文本数据。为了构建鱼类知识子图的模式图，本文利用 HTML 封装器从 fishdb 和 FishBase 中抽取概念和上下位关系，从百科页面中抽取概念的属性，并利用多策略学习方法从以上数据源中迭代地抽取同义词关系。在数据图的构建上，本文从 fishdb 和 FishBase 中抽取鱼类实例，采用多种方法获取实例的属性值。例如，使用 HTML 封装器从心食谱网站中获取属性“鱼类美食”的值，使用模式从《中国食物成分表》(2002 年版)中获取属性“营养成分”的值。

海洋知识图谱的构建结果经过海洋知识专家检查并处理数据冲突后发布，目前包含了全球已命名的 3 万余种鱼类和 20 多个属性，提供海洋知识可视化、语义知识检索、海洋知识推荐等知识服务。海洋知识图谱提供车轮视图、树状视图和详情视图 3 种可视化检索方式，分别侧重展示实体间的语义关系、海洋知识图谱的体系结构以及实体、概念的属性详情。此外，其

提供的语义搜索服务，在为用户输入的自然语言问题提供直接答案的同时，还展示实体的知识卡片和相关实体，并结合图书馆的资源返回文献搜索结果。如图 5 所示，输入的问题“小黄鱼的分布”被解析出实体“小黄鱼”和语义“小黄鱼的分布生态系统”，基于此，系统返回语义检索结果、“小黄鱼”的知识卡片和相关实体，以及相关的文献资源。



图 5　海洋知识图谱的语义搜索展示

5.3　企业知识图谱

企业知识图谱数据整合了 3 000 万家企业数据以及来自互联网的专利数据与招投标数据。首先，领域专家构建了行业知识图谱模式图，包含人物、公司、股票、专利、投资和招标等顶层概念。再者，利用 D2R 工具将企业提供的基于关系数据库的企业信息转化成 RDF（Resource Definition Framework）数据，构成了基础的企业知识图谱。但此时的企业知识比较简单，需要通过其他数据来源的数据进行补充。本文先增加专利与招投标信息：从中国政府采购网、中国专利信息网等网站抓取文本公告，基于启发式信息定义模式抽取企业招投标信息和专利信息。然后基于百科与新闻进一步补充企业信息，包括高管信息的变动，企业兼并与收购信息等。

企业知识图谱可以提供实际控制人查询与关系发现等功能。其中，实际控制人查询功能是指查询对企业占股最大的自然人，由于个人对企业的控制可以是直接投资，也可以通过个人控制的企业再来投资企业，因此，算法基于图的遍历算法实现。用户输入一个企业，系统可以返回该企业的实际控制人。关系发现功能可以发现公司或人物之间的间接关系，图 6 展示了“中国铝业股份有限公司”和“中信证券股份有限公司”之间的关系，其中，箭头代表投资关系，该图说明，“中国铝业股份有限公司”和“中信证券股份有限公司”的投资方，共同投资了企业 B。

知识管理论坛，2016(3):226–234
DOI：10.13266/j.issn.2095–5472.2016.027

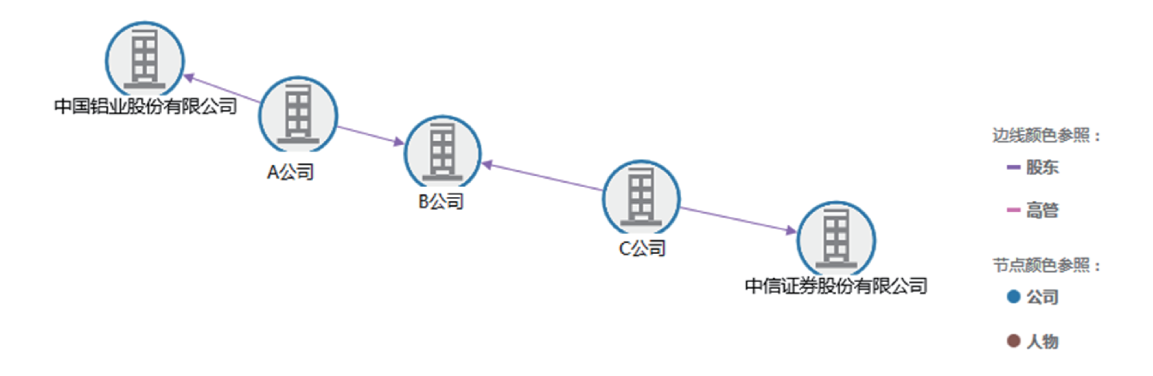


图 6 企业知识图谱中的企业关系发现

6 结语

本文对知识图谱进行了形式化定义，并详细描述了数据驱动的增量式知识图谱构建方法。以该方法分别构建了中医药知识图谱、海洋知识图谱和企业知识图谱，并开发了相关应用。以上 3 个垂直知识图谱的构建证明了本文提出的构建方法的有效性，体现了图谱在知识融合方面的优势；它们的相关应用反映了知识图谱在不同领域的应用价值。

参考文献：

[1] SINGHAL A. Introducing the knowledge graph: things, not strings[EB/OL]. [2012-05-16]. https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html.

[2] BIEGA J, KUZEY E, SUCHANEK F M. Inside YAGO2s: a transparent information extraction architecture[C]//Proceedings of the 22nd international conference on World Wide Web companion. Rio de Janeiro: International World Wide Web Conferences Steering Committee, 2013: 325-328.

[3] BIZER C, LEHMANN J, KOBILAROV G, et al. DBpedia-A crystallization point for the Web of data[J]. Web Semantics: science, services and agents on the World Wide Web, 2009, 7(3): 154-165.

[4] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//Proceedings of the 2008 ACM SIGMOD international conference on management of data. Vancouver: ACM, 2008: 1247-1250.

[5] CARLSON A, BETTERIDGE J, KISIEL B, et al. Toward an architecture for never-ending language learning[C]//Proceedings of the twenty-fourth AAAI Conference on artificial intelligence. Atlanta: AAAI Press, 2010: 3.

[6] NIU X, SUN X, WANG H, et al. Zhishi. me-weaving chinese linking open data[M]. The Semantic Web–ISWC 2011. Berlin: Springer, 2011: 205-220.

[7] HU F, SHAO Z, RUAN T. Self-supervised Chinese ontology learning from online encyclopedias[J]. The scientific world journal, 2014, 2(11):1-13.

[8] BRICKLEY D, GUHA R V.RDF vocabulary description language 1.0: RDF schema[EB/OL]. [2003-12-15]. https://www.w3.org/TR/2003/PR-rdf-schema-20031215/.

[9] BECHHOFER S, VAN HARMELEN F, HENDLEr J, et al. OWL web ontology language reference, 2004[EB/OL]. [2004-02-10]. https://www.w3.org/TR/owl-ref/.

[10] PRUDHOMMEAUX E, SEABORNE A. SPARQL query language for RDF[EB/OL]. [2008-01-15]. https://www.w3.org/TR/rdf-sparql-query/.

[11] MILLER G A. WordNet: a lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.

[12] RAMACHANDRAN D, REAGAN P, GOOLSBEY K. First-orderizedresearchcyc: expressivity and efficiency in a common-sense ontology[C]//AAAI workshop on contexts and ontologies: theory, practice and applications. Pittsburgh: AAAI Press, 2005: 25-34.

[13] RUAN T, DONG X, WANG H, et al. Evaluating and comparing web-scale extracted knowledge bases in Chinese and English[C]//Joint international semantic technology conference. Yichang: Springer International Publishing, 2015: 167-184.

[14] SCHMITZ M, BART R, SODERLAND S, et al. Open language learning for information extraction[C]//Proceedings



of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning. Jeju Island: Association for Computational Linguistics, 2012: 523-534.

[15] RUAN T, LIN Y, WANG H, et al. A multi-strategy learning approach to competitor identification[M]//Semantic Technology. Berlin: Springer International Publishing, 2014: 197-212.

[16] MINTZ M, BILLS S, SNOW R, et al. Distant supervision for relation extraction without labeled data[C]//Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP: Volume 2. Singapore: Association for Computational Linguistics, 2009: 1003-1011.

[17] WANG H, FANG Z, ZHANG L, et al. Effective online knowledge graph fusion[M]. The Semantic Web-ISWC 2015. Berlin: Springer International Publishing, 2015: 286-302.

[18] PROCTOR M. Drools: a rule engine for complex event processing[M]//Applications of graph transformations with industrial relevance. Berlin: Springer, 2011: 2.

[19] CYGANIAK R, BIZER C, GARBERS J, et al. The d2rq mapping language[EB/OL]. [2007-08-24]. http://d2rq.org/.

[20] SHAO K T. Taiwan fish database[EB/OL]. [2014-11-10]. http://fishdb.sinica.edu.tw.

[21] FROESE R, PAULY D. Fishbase[EB/OL]. [2012-06-15]. http://www.fishbase.org.

Research on the Construction and Applicationof Vertical Knowledge Graphs

Ruan Tong[1] Wang Mengjie[2] Wang Haofen[1] Hu Fanghuai[3]

[1]School of Information Science and Engineering of East China University of Science and Technology, Shanghai 200237

[2]Department of Computer Science & Engineering of East China University of Science and Technology, Shanghai 200237

[3]Shanghai Hi Knowledge Technology Co. Ltd., Shanghai 200433

Abstract: [Purpose/significance] In recent years, knowledge graphs have gained wide attention from academia and industry. Knowledge graphs of different domains are widely used in query understanding, automatic question answering and document representation. **[Method/process]** In this paper, the construction of vertical knowledge graphs and related applications were studied. Specifically, we first gave a formal definition of knowledge graph. Then a data-driven incremental constructing method was proposed. We put emphasis on constructing a data graph of a vertical knowledge graph. Based on the proposed method, we have built Traditional Chinese Medicine Knowledge Graph, Marine-oriented Knowledge Graph and Security-oriented Knowledge Graph. **[Result/conclusion]** These vertical knowledge graphs illustrate the feasibility of our method and extensive usability of vertical knowledge graphs.

Keywords: knowledge graph knowledge acquisition knowledge fusion semantic search prescription assistance relation discovery